

Economics Department Discussion Papers Series

ISSN 1473 – 3307

A Randomized Procedure for Choosing Data Transformation

Valentina Corradi and Norman R. Swanson

Paper number 01/05

A Randomized Procedure for Choosing Data Transformation[¤]

Valentina Corradi
University of Exeter, U.K.

Norman R. Swanson
Texas A&M University, U.S.A.

June 2001

JEL classification: C12, C22.

Keywords: Deterministic trend, nonlinear transformation, nonstationarity, randomized procedure.

[¤]Valentina Corradi, Department of Economics, University of Exeter, Streatham Court, Exeter EX4 4PU, U.K., v.corradi@exeter.ac.uk. Norman R. Swanson, Department of Economics, Texas A&M University, College Station, Texas 77843-4228, USA. The authors wish to thank Herman Bierens, Graham Elliot, Liudas Giraitis, Eric Ghysels, Clive Granger, Vassilis Hajivassiliou, Javier Hidalgo, Nick Keifer, Shinichi Sakata, Frank Schorfheide, Yongcheol Shin, Andy Snell, and seminar participants at Arizona State University, Indiana University, London School of Economics, Ohio State University, Pennsylvania State University, University of Edinburgh, Queen Mary and Westfield College, Warwick University, Texas A&M University, the 1999 Summer Meeting of the Econometric Society, the 1999 U.K. Econometric Group Meeting in Bristol, the 1999 Rotterdam Meeting on Nonlinear Modelling of Multivariate Macroeconomic Relations, and the 1999 Svinklov Meeting on Nonlinear Modelling in Economics for helpful comments on an earlier version of this paper.

Abstract

Standard unit root and stationarity tests (see e.g. Dickey and Fuller (1979)) assume linearity under both the null and the alternative hypothesis. Violation of this linearity assumption can result in severe size and power distortion, both in finite and large samples. Thus, it is reasonable to address the problem of data transformation before running a unit root test. In this paper we propose a simple randomized procedure, coupled with sample conditioning, for choosing between levels and log-levels specifications in the presence of deterministic and/or stochastic trends. In particular, we add a randomized component to a basic test statistic, proceed by conditioning on the sample, and show that for all samples except a set of measure zero, the statistic has a \hat{A}^2 limiting distribution under the null hypothesis (log linearity), while it diverges under the alternative hypothesis (level linearity). Once we have chosen the proper data transformation, we remain with the standard problem of testing for a unit root, either in levels or in logs. Monte Carlo findings suggest that the proposed test has good finite sample properties for samples of at least 300 observations. In addition, an examination of the King, Plosser, Stock and Watson (1991) data set is carried out, and evidence in favor of using logged data is provided.

1 Introduction

In empirical economics, unit root tests are typically performed using logs. This is consistent with much of the real business cycle literature (see e.g. Long and Plosser (1993) and King, Plosser, Stock, and Watson (1991)) where it is suggested, for example, that GDP should be modeled in logs, given an assumption that output is generated according to a Cobb-Douglas production function. While this may be convenient from a statistical perspective¹, there is no clear theoretical reason why logs should be used rather than levels, when performing unit root tests, for example. This distinction is important because standard unit root tests assume linearity under both the null and the alternative hypothesis, and violation of this linearity assumption can result in severe size and power distortion, both in finite and large samples (see e.g. Granger and Hallman (1991)). Furthermore, correct choice of data transformation is crucial when specifying forecasting models using integrated and/or cointegrated variables, as documented in Arino and Franses (2000) and in Chao, Corradi and Swanson (2001), for example.

The current convention is to define an integrated process of order \mathbf{d} (say $I(\mathbf{d})$, using the terminology of Engle and Granger (1987)) as one which has the property that the partial sum of the \mathbf{d} th difference, scaled by $T^{1-\mathbf{d}}$, satisfies a functional central limit theorem (FCLT). Under this definition, integratedness in logs does not imply integratedness in levels, or **vice versa**. Thus, any **a priori** assumption concerning whether to model data in levels or logs has important implications for the outcome of unit root and related tests, as the limiting behavior of these tests is generally examined using FCLTs. For example, Granger and Hallman (1991) show that the percentiles of the empirical distribution of the Dickey-Fuller (1979) statistic constructed using $\exp(\mathbf{X}_t)$ are much higher, in absolute value, than the corresponding percentiles constructed using the original time series \mathbf{X}_t , when \mathbf{X}_t is a random walk process. Thus, inference based on the Dickey-Fuller statistic using the exponential transformation leads to an over-rejection of the unit root null hypothesis, when standard critical values are used. More recently, it has been shown in Corradi (1995) that if \mathbf{X}_t is a random walk, then any convex transformation (such as exponentiation) is a submartingale, and any concave transformation (such as taking logs) is a supermartingale. However, while submartingales and supermartingales have a unit root component, their first differences do not

¹The Cobb-Douglas production function is linear in the variables after logs have been taken, so that there is arguably an element of convenience involved in the choice to take logs of the data in this and related contexts.

generally satisfy typical FCLTs. Thus, Dickey-Fuller type tests no longer have well defined limiting distributions. Given all of the above considerations, it is of some interest to use a statistical procedure for selecting between linear and loglinear specifications, rather than simply assuming from the outset that a series is best modeled as either linear or loglinear. This is particularly true when unit root tests are to be subsequently carried out, as the researcher must then choose between logs and levels prior to carrying out the tests.

In recent years, the choice of data transformation for nonstationary series (henceforth, by nonstationary we mean $I(1)$) received considerable attention. Important contributions in the area include De Bruin and Franses (1999), Franses and Koop (1998), Franses and McAleer (1998), Kobayashi (1994) and Kobayashi and McAleer (1999a,b). One line of research (see e.g. Franses and Koop (1998) and Franses and McAleer (1998)) analyzes the joint problem of choosing the Box-Cox transformation (with levels and logs being special cases) and choosing between stationarity and nonstationarity. Monte Carlo results reported in the papers just cited are rather encouraging, and suggest that the problem can be treated as discussed by the authors. However, work still remains to be done before a complete picture of the asymptotic behavior of such tests based on Box-Cox transformations can be obtained. Broadly speaking, the main issue that arises when studying the limiting behavior of these and related tests (e.g. tests constructed under both nonstationarity and nonlinearity) can be summarized as follows. Often, test statistics can be written in “ratio” form, where the denominator of the test is an estimator of a (long run) variance. In such cases, a well defined limiting distribution can be derived under the null hypothesis. However, under the alternative hypothesis, it is often the case that both the numerator and the denominator approach infinity, with the latter diverging at a faster (or at least not slower) rate than the former. As a consequence, some tests have zero asymptotic power against alternatives of interest. This problem is solved in a rather ingenious way in a recent paper by Kobayashi and McAleer (KM: 1999a), who propose a test for distinguishing between levels and logs in models with a unit root. In particular, by assuming that the variance of the innovation process approaches zero at a sufficiently fast rate as the sample increases, KM derive the limiting distribution of their test under the null hypothesis, and show that the probability of type II error approaches zero asymptotically.² KM stress that their innovation process assumption is not only plausible, but also necessary whenever the objective is to ensure

²The device that KM use is called small sigma asymptotics (see e.g. Bickel and Doksum (1981)).

positivity for linear $I(1)$ DGPs with no deterministic trend. The current paper is meant to continue the line of research begun by KM.

Our objective in this paper is to propose a procedure for distinguishing between the null hypothesis of a loglinear DGP and a (level) linear DGP. Once we have chosen the correct data transformation, we can proceed by testing for $I(0)$ versus $I(1)$ using standard unit root and stationarity tests. Two points are worth making at this juncture. First, when defining the relevant models from among which to choose, we allow for rather general, dependent error processes. Thus, the test is robust to a rich variety of dynamics. Second, we overcome the test consistency problem discussed above by basing our test on the combined use of a randomization procedure coupled with sample conditioning. In particular, we add randomness to our basic statistic, proceed by conditioning on the sample, and show that for all samples except a set of measure zero, the statistic has a chi-squared limiting distribution under the null hypothesis, while it diverges under the alternative hypothesis. Clearly, then, the asymptotic behavior of the statistic is driven by the probability measure governing the added randomness. As is typical when using randomization procedures, different users, using the same sample, may obtain different statistic values. Nevertheless, conditional on the sample and for all samples except a set of measure zero, we choose the null hypothesis with probability approaching one whenever it is true, and we reject the null hypothesis with probability approaching one whenever it is false.

In a series of Monte Carlo experiments, we establish that the finite sample properties of the test are quite good for samples of at least 300 observations, for DGPs calibrated using U.S. monetary data. In addition, an empirical illustration is provided in which the King, Plosser, Stock and Watson (1991) data set is examined. Results suggest that many macroeconomic variables are “best” modelled as loglinear.

The rest of the paper is organized as follows. Section 2.1 discusses problems involved with testing for unit roots in the presence of incorrect data transformation, and introduces the randomized statistic. Section 2.2 studies the asymptotic behavior of the proposed statistic. The findings from a Monte Carlo exercise are reported in Section 3, and a small empirical illustration is given in Section 4. Concluding remarks are gathered in Section 5. All proofs are collected in an appendix. Hereafter, $\mathcal{Q}^n_{a:s} \rightarrow \mathcal{Q}$ denotes convergence in distribution conditional on the sample, s (i.e. for all sample except a set of measure zero).

2 Distinguishing Between I(0) and I(1) Processes in Logs and Levels

2.1 Set Up

Given a series of observations on an underlying strictly positive process, \mathbf{X}_t , $t = 1; 2; \dots$, our objective is to decide whether: (1) \mathbf{X}_t is an I(0) process around a linear deterministic trend, (2) $\log \mathbf{X}_t$ is an I(0) process possibly around a nonzero linear deterministic trend, (3) \mathbf{X}_t is an I(1) process around a positive linear deterministic trend, and (4) $\log \mathbf{X}_t$ is an I(1) process, possibly around a linear deterministic trend. More precisely we want to choose among the following DGPs:

$$H_1 : \mathbf{X}_t = \mu_0 + \alpha_0 t + \frac{1}{2} \mathbf{X}_{t-1} + \varepsilon_{1,t}; j\frac{1}{2} < 1 \text{ and } \alpha_0 > 0;$$

$$H_2 : \mathbf{X}_t = \mu_0 + \mathbf{X}_{t-1} + \varepsilon_{1,t}; \alpha_0 > 0;$$

$$H_3 : \log \mathbf{X}_t = \mu_1 + \alpha_1 t + \frac{1}{2} \log \mathbf{X}_{t-1} + \varepsilon_{2,t}; j\frac{1}{2} < 1 \text{ and } \alpha_1 \geq 0 \text{ and}$$

$$H_4 : \log \mathbf{X}_t = \mu_1 + \log \mathbf{X}_{t-1} + \varepsilon_{2,t}; \alpha_1 \geq 0;$$

Note that in order to ensure positivity we assume that the DGPs in levels have a positive trend component. As discussed above, Kobayashi and McAleer (1999a) employ small sigma asymptotics for showing the limiting behavior of their statistic in the case where $\alpha_i = 0$; $i = 1; 0$. As they stress, this device is in general suitable for financial series, for example. Our test is complementary to theirs as we propose using a different device (randomization), which may be useful when looking at macroeconomic data, for example.

While it is easy to define a test that has a well defined distribution under one of H_1 to H_4 , it is not clear how to ensure that the test has power against all of the remaining DGPs. To illustrate the problem, consider the sequence, \mathbf{z}_t , given as the residuals from a regression of \mathbf{X}_t on a constant and a time trend. In particular, construct the statistic for the null of stationarity proposed by Kwiatkowski, Phillips, Schmidt, and Shin (KPSS: 1992):

$$S_T = \frac{1}{\hat{\gamma}_T^2} T^{-1/2} \sum_{t=1}^T \bar{\mathbf{X}}_t' \mathbf{X}_t \mathbf{z}_t^2 \mathbf{A}_T^{-1/2};$$

where $\hat{\gamma}_T^2$ is a heteroskedasticity and autocorrelation (HAC) robust estimator of $\text{var} \left(T^{-1/2} \sum_{j=1}^3 \mathbf{P}_{j=1}^T \mathbf{z}_t^2 \right)$. It is known from KPSS that if \mathbf{X}_t is I(0) (possibly around a linear deterministic trend), then S_T has a well defined limiting distribution under the null hypothesis, while S_T diverges at rate $T^{-1/2}$ under the alternative that \mathbf{X}_t is an I(1) process, where l_T is the lag truncation parameter used in the es-

timination of the variance term in S_T . However, if the underlying DGP is $\log X_t = \pm_1 + \log X_{t-1} + \varepsilon_t$, $\pm_1 > 0$ (i.e. $\log X_t$ is a unit root process) then both $\frac{1}{T} \sum_{t=1}^T \varepsilon_t^2$ and $\frac{1}{T} \sum_{t=1}^T \left(\sum_{j=1}^t \varepsilon_j \right)^2$ will tend to diverge at a geometric rate, given that $X_t = \exp(\log X_0 + \pm_1 t + \sum_{j=1}^t \varepsilon_j)$. In this case it is not clear whether the numerator or the denominator is exploding at a faster rate. This problem is typical of all tests which are based on functionals of partial sums and variance estimators, and arises because certain **nonlinear** alternatives are not treatable using standard FCLTs.

Recently Park and Phillips (PP: 1999, 2001) have developed an asymptotic theory for partial sums and for moments of nonlinear functions of integrated processes. The novel and important approach of Park and Phillips is based on the idea of replacing sample sums by spatial sums and then analyzing the average time spent by the process in the vicinity of given points. A key ingredient is the notion of local time of a Brownian motion. In our setup, we need to take into account the presence of a positive deterministic trend, at least for levels DGPs, however, and we are currently unable to generalize the PP results to the case of processes with deterministic drift components. The intuition behind the difficulty in providing such a generalization stems from the fact that we cannot embed an integrated process with deterministic drift into a continuous semimartingale³, and to the best of our knowledge a local time theory is available only for continuous semimartingale processes. Broadly speaking, an integrated process with positive drift is dominated by the deterministic component and so it is transient. Thus, compact sets in the state space will be visited only a finite number of times, as the process will spend almost all time in the “proximity of infinity”. Therefore, we shall follow a different approach, based on the combination of randomization and sample conditioning. In the sequel, in order to distinguish between $H_1; H_2; H_3$ and H_4 above, we rely on the following assumption:

Assumption A1: (i) $X_t > 0; \forall t \geq 0$; (ii) $\varepsilon_{i,t}; i = 1, 2$ is a zero-mean strictly stationary strong mixing process with mixing coefficient α_m satisfying $\sum_{m=0}^{\infty} \alpha_m^{\frac{1}{4+2\delta}} < \infty$; for any $\delta > 0$; and (iii) $0 < E(\varepsilon_{i,1}^2) = \sigma_i^2 < \infty$ and $E(\varepsilon_{i,1}^2 \varepsilon_{j,1}^2) < \infty$; $i = 1, 2$; for the same δ as in (ii).

Note that Assumption A1 suffices for the partial sums of $\varepsilon_{j,t}$ to satisfy a strong (and so a weak) invariance principle (see e.g. Corollary 4.1 and Theorem 3.1 in Berger (1990)).⁴ As mentioned

³A semimartingale is a process given by the sum of a martingale plus an adaptive process of finite variation (see e.g. Revuz and Yor (1990), pp.121).

⁴The strict stationarity assumption can be relaxed at the price of strengthening the mixing condition. In fact, a strong invariance principle for strong mixing, non-stationary processes could be used (see e.g. Theorem 2 in Eberlain (1986)).

above, our main objective is to distinguish between levels and logs. This is because once we have chosen the correct data transformation, we can choose between $I(0)$ and $I(1)$ via standard tests.

Now, group the above hypotheses as follows:

$$H_0 : H_3 \text{ [} H_4; \pm_1 > 0$$

$$H_A : H_1 \text{ [} H_2$$

Thus, the null hypothesis is logs and the alternative is levels. The case of $\pm_1 = 0$; (i.e. no deterministic drift in the log DGPs) is somewhat more complex and will be treated subsequently.

The proposed test statistic is:

$$S_{T;R}(!) = \int_U Z_{T;R}^2(u; !)\mathbb{W}(u)du; \quad (1)$$

where U is a compact set on the real line, $!$ denotes the dependence of $S_{T;R}(!)$ on the data, $\int_U \mathbb{W}(u)du = 1$,

$$Z_{T;R}(u) = \frac{1}{\sqrt{R}} \sum_{i=1}^n (1 - V_{\gg;T}^i(!) \cdot u^{\frac{1}{2}}); \quad (2)$$

with $R = o(T)$; and $V_{\gg;T}^i(!)$ is defined as:

$$V_{\gg;T}^i(!) = \frac{\bar{A}}{T} \sum_{t=1}^T \mu \frac{\Delta X_t}{\Delta X_1} \mathbb{W}_{1=2}^{!} \gg_i; \quad i = 1; \dots; R; \quad (3)$$

where \gg_i is an $\text{iidN}(0;1)$ random variable. Note that we divide all data by the initial value in order to make the statistic invariant to scalar multiplication of the observations. It turns out that for any sample, $!$; which is a realization of a DGP under the null hypothesis (i.e. a log DGP), $S_{T;R}(!)$ converges in distribution to a \hat{A}^2 random variable, while for any $!$ which is a realization of a DGP under the alternative hypothesis (i.e. a level DGP), $S_{T;R}(!)$ diverges. Note that, as we proceed conditionally on the sample, the asymptotic behavior of the statistic is driven by the probability law governing the artificial randomness (i.e. the probability law governing \gg_i): Randomized procedures have previously been used in the literature. For example, Dufour and Kiviet (1996) use a randomized test to obtain finite sample confidence intervals for structural changes in dynamic models; although in finite samples the level of the actual and of the randomized test may differ, they are equivalent in large samples. In a different context, Lütkepohl and Burda (1997) use a randomized approach for constructing Wald tests under non regular conditions - namely when the matrix of partial derivatives has reduced rank. They essentially overcome a certain singularity

problem by adding randomness, and convergence to the limiting distribution is driven by both the probability law governing the sample and the probability law governing the added randomness. What differentiates our approach from the randomized procedures cited above is the joint use of randomization and sample conditioning. Our asymptotic result only holds conditionally on the sample, and for all samples except a set of measure zero. It is also worth noting, however, that randomization coupled with sample conditioning is used elsewhere to obtain conditional p-values and conditional percentiles, for example, when the limiting distribution of the actual statistic is data dependent (see e.g. Hansen (1996), Corradi and Swanson (2001) and Inoue (2001)). In these cases, though, inference is based on comparison of the actual statistic (which depends only on the sample) with conditional percentiles. In the present context, inference is based on the randomized statistic, conditional on the sample.

3 Asymptotic Results

Hereafter let \mathbf{d}^n denote convergence in distribution according to \mathbf{P}^n ; the probability law governing \mathbf{y}_i ; $i = 1, \dots, R$; conditional on the sample. Also, \mathbf{E}^n and \mathbf{Var}^n denote the mean and the variance operators with respect to the probability law \mathbf{P}^n . Finally, the notation $\mathbf{a.s.} \mathbf{; } \mathbf{j} \mathbf{!}$ means conditional on the sample, and for all samples except a set of measure zero.

Theorem 1: Let A1 hold. If $R = T^a$; $0 < a < 1$; then as $T \rightarrow \infty$:

- (i) Under H_0 , $S_{T;R}(\mathbf{!}) \xrightarrow{d^n} \hat{A}_1^2$; $\mathbf{a.s.} \mathbf{; } \mathbf{j} \mathbf{!}$;
- (ii) Under H_A , there exists a $\delta^0 > 0$ such that $\delta^0 < 1$; $\mathbf{P}^n \left(\frac{1}{R} S_{T;R}(\mathbf{!}) > \delta^0 \right) \rightarrow 1$, $\mathbf{a.s.} \mathbf{; } \mathbf{j} \mathbf{!}$;

Thus, the test statistic has a well defined limiting distribution for each sample which is a realization of a DGP under H_0 and diverges for each sample which is a realization of a DGP under H_A .

It is worth noting that the interpretation of **test size** in the current context differs from the interpretation associated with inference which is not sample conditioned. To see this difference, consider the following example. Suppose we draw 10000 samples from a DGP generated under H_0 . In addition, there are 10000 people performing the same test. According to the usual definition, the size is 5% if all 10000 people decide in favor of H_0 based on examination of 9500 samples, while they all decide in favor of H_A based on the remaining 500 samples. On the other hand, for the

sample conditioned statistic, some group⁵ of 9500 people decide in favor of \mathbf{H}_0 for each of the 10000 samples, while the remaining 500 people decide in favor of the alternative for each sample.

Although a detailed proof of the theorem above is given in the appendix, it is perhaps worthwhile to give an intuitive explanation of the result. Note first that conditional on the sample, $\mathbf{V}_{\gg;T}^i(!) \gg \mathbf{N}(0; \frac{1}{T} \sum_{t=1}^T \frac{\phi X_t}{\phi X_1})^2$. Now, note that under the null hypothesis of a log DGP, $\frac{1}{T} \sum_{t=1}^T \frac{\phi X_t}{\phi X_1}$ diverges to infinity at a geometric rate as T gets large. It then follows that $\mathbf{V}_{\gg;T}^i(!)$ diverges almost surely to $+\infty$ or to $-\infty$, a.s. In addition, because of symmetry we have that $\mathbf{V}_{\gg;T}^i(!)$ diverges to either plus or minus infinity with probability approaching $1/2$ a.s. Thus, $E(\mathbf{1}_{\mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u} > 0}) = P(\mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u} > 0) = \frac{1}{2} + o(1)$; uniformly in \mathbf{u} for \mathbf{U} compact, and $\text{Var} \frac{1}{R} \sum_{i=1}^R \mathbf{1}_{\mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u} > 0} = \frac{1}{4} + o(1)$, uniformly in \mathbf{u} a.s. The desired result then follows directly from the central limit theorem for independent triangular arrays.⁶ Under the alternative hypothesis of a level DGP, by the strong law of large numbers, $\frac{1}{T} \sum_{t=1}^T \frac{\phi X_t(!)}{\phi X_1(!)}$ converges almost surely to a constant, say \mathbf{M} . Let $\mathbf{F}(\mathbf{u})$ be the CDF of a $\mathbf{N}(0; \mathbf{M})$ random variable, evaluated at \mathbf{u} : Now,

$$\frac{1}{R} \sum_{i=1}^R (\mathbf{1}_{\mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u} > \frac{1}{2}}) = \frac{1}{R} \sum_{i=1}^R (\mathbf{1}_{\mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u} > \mathbf{F}(\mathbf{u})}) + \frac{1}{R} (\mathbf{F}(\mathbf{u}) - \frac{1}{2}): \quad (4)$$

The first term on the right hand side above is bounded in probability, because of the central limit theorem for empirical processes for independent triangular arrays, while the second term diverges at rate $\frac{1}{R}$ whenever $\mathbf{F}(\mathbf{u}) \neq \frac{1}{2}$ (i.e. whenever $\mathbf{u} \neq 0$):

In practice, the interval over which \mathbf{u} is integrated must be determined. For increasing width intervals which are centered at zero and for \mathbf{u} uniform over \mathbf{U} , finite sample power improves, while finite sample size deteriorates. The dependence of finite sample power on \mathbf{U} in this case can be seen immediately from equation (4), as the second term on the right hand increases the further \mathbf{u} is from zero. On the other hand, finite sample size tends to get worse the larger is \mathbf{u} : Hence, there is a trade-off between finite sample size and power associated with the choice of the interval \mathbf{U} . In practice, we also have to choose \mathbf{R} : It is easy to see that the higher is the rate at which \mathbf{R} grows, provided it grows at a slower rate than \mathbf{T} ; the higher is the finite sample power. The choice of \mathbf{U} and \mathbf{R} is analyzed in the Monte Carlo section below.

⁵The members of the group may change from sample to sample.

⁶Note that conditionally on sample, $\mathbf{V}_{\gg;T}^i$, $i = 1, \dots, R$ is an independent triangular array.

We now turn to the case where $\pm_1 = 0$ (i.e. the case of log DGPs without a deterministic trend component). For example, under H_4 , $\Delta X_t = X_{t-1} \exp(\epsilon_{2,t} - 1)$; where $X_{t-1} = \exp(\log X_0 + \sum_{j=1}^{t-1} \epsilon_{2,j})$. As $\sum_{j=1}^T \epsilon_{2,j}$ diverges either to plus or minus infinity, it follows that $\frac{1}{T} \sum_{t=1}^T \frac{\epsilon_{2,t}^2}{\epsilon_{2,1}^2}$ either diverges to infinity or converges to zero. Thus, $V_{\gg,T}^i(!)$ either diverges to ∞ or converges to zero, depending on $!$. Intuitively, if $V_{\gg,T}^i(!)$ converges to zero; $1fV_{\gg,T}^i(!) \rightarrow u g \rightarrow 1$; for all $u > 0$; and $1fV_{\gg,T}^i(!) \rightarrow u g \rightarrow 0$; for all $u < 0$. On the other hand, when $V_{\gg,T}^i(!)$ diverges to ∞ ; $1fV_{\gg,T}^i(!) \rightarrow u g \rightarrow 1$ (resp. 0) with probability $\frac{1}{2}$; a.s.; for all $u \in U$; U compact. Needless to say, it is unknown whether $\frac{1}{T} \sum_{t=1}^T \frac{\epsilon_{2,t}^2}{\epsilon_{2,1}^2}$ converges to zero or diverges, for any given sample. A natural approach is thus to construct two statistics and then base inference on the smaller one. Without loss of generality, let U^+ be a compact set on the positive real line (including 0)⁷. Define:

$$S_{T;R}^a(!) = \int_{U^+} \left(\frac{1}{R} \sum_{i=1}^R (1 - V_{\gg,T}^i(!) \cdot u^{\frac{1}{2}})^2 \right) \mathbb{1}(u) du$$

and

$$S_{T;R}^b(!) = \int_{U^+} \left(\frac{1}{R} \sum_{i=1}^R (1 - V_{\gg,T}^i(!) \cdot u^{\frac{1}{p}})^2 \right) \mathbb{1}(u) du; \quad p = 1;$$

where $\int_{U^+} \mathbb{1}(u) du = 1$. Note that $S_{T;R}^a(!)$ is the same as $S_{T;R}(!)$ above, with the additional requirement that it is computed over U^+ . The choice between logs and levels in this context is facilitated by using $\min(S_{T;R}^a(!); S_{T;R}^b(!))$. The intuition for this test is as follows. Conditioning on a sample for which $\frac{1}{T} \sum_{t=1}^T \frac{\epsilon_{2,t}^2}{\epsilon_{2,1}^2} \rightarrow 1$ implies that $S_{T;R}^a(!)$ is asymptotically \hat{A}_1^2 ; while $S_{T;R}^b(!)$ diverges. On the other hand, conditioning on a sample for which $\frac{1}{T} \sum_{t=1}^T \frac{\epsilon_{2,t}^2}{\epsilon_{2,1}^2} \rightarrow 0$; implies that $S_{T;R}^a(!)$ diverges; while $S_{T;R}^b(!)$ converges in probability to zero. This suggests using the above test within the context of the following hypotheses,

$H_0^0 : H_4$ with $\pm_1 = 0$ and

$H_A^0 : H_1 \cup H_2 \cup H_3$ with $\pm_1 = 0$:

Theorem 2: Let Assumption A1 hold. If $R = T^a$; $0 < a < 1$; then as $T; R \rightarrow \infty$:

- (i) Under H_0^0 ; $\lim_{T;R \rightarrow \infty} P^{\frac{1}{R}} \min(S_{T;R}^a(!); S_{T;R}^b(!)) > c^-$; a.s.; where c^- is the $(1 - \frac{1}{R})$ th percentile of a \hat{A}_1^2 random variable.
- (ii) If in addition, $E(\exp(\epsilon_{1,t}^{2(2+\hat{A})})) < \infty$; for some $\hat{A} > 0$; then under H_A^0 there exists $\delta > 0$; such that $\lim_{T;R \rightarrow \infty} P^{\frac{1}{R}} \min(S_{T;R}^a(!); S_{T;R}^b(!)) > \delta > 0$; a.s.:

⁷Analogously, for U^- a compact set on the negative real line, set p in $S_{T;R}^b(!)$ equal to zero. Theorem 2 then holds for the min statistic defined on U^- .

Thus, the asymptotic type I error is less than or equal to α , while the asymptotic type II error is zero, conditional on \mathcal{H}_1 , and for all \mathcal{H}_1 except a set of measure zero. Theorem 2 also holds for $\alpha_1 > 0$. In this case, the smaller statistic is $S_{T;R}^a(\alpha)$; \mathcal{H}_1 : In the case where the test selects \mathcal{H}_A^0 one cannot distinguish between DGPs in levels with short memory (i.e. stationarity) and DGPs in logs (i.e. \mathcal{H}_A^0 contains both \mathcal{H}_1 and \mathcal{H}_3). In this case, it remains only to test the significance of the coefficient on a linear deterministic trend in a levels regression. Even if the process is actually short memory in logs, the test is well defined, as the exponential of a short memory process is short memory. Thus, a finding that the coefficient on the trend component is significant implies the consequent choice of levels data, otherwise use logged data. In the previous section, it was noted that a larger compact set, \mathcal{U} , leads to higher finite sample power as well as higher finite sample size, for \mathcal{U} centered around zero. In the current context, finite test performance trade-offs are not as straightforward. Consider $\mathcal{U}^+ = [0; \mathbf{u}_{\max}]$: For all samples in which $\frac{1}{T} \sum_{t=1}^T \frac{\phi X_t}{\phi X_1}^2$ converges to zero, larger \mathbf{u}_{\max} implies better finite sample size and worse finite sample power. On the other hand, the opposite holds for all samples in which $\frac{1}{T} \sum_{t=1}^T \frac{\phi X_t}{\phi X_1}^2$ diverges. For this reason, we recommend use of the statistic which is defined for \mathcal{U} (see Theorem 1). If the null hypothesis is rejected, but there is ancillary evidence that the true DGP may be a unit root process in logs with no drift, continue by using the statistic described in Theorem 2.

4 Finite Sample Evidence

In this section the results of a small set of Monte Carlo experiments are reported. Data are generated according to \mathcal{H}_1 to \mathcal{H}_4 in Section 2.1, which are here written as,

$$\mathcal{H}_1 : X_t = \alpha_1 + \alpha_1 t + \frac{1}{2} X_{t-1} + \epsilon_{1,t};$$

$$\mathcal{H}_2 : X_t = \alpha_2 + X_{t-1} + \epsilon_{2,t};$$

$$\mathcal{H}_3 : \log X_t = \alpha_3 + \alpha_3 t + \frac{1}{2} \log X_{t-1} + \epsilon_{3,t},$$

$$\mathcal{H}_4 : \log X_t = \alpha_4 + \log X_{t-1} + \epsilon_{4,t};$$

where all errors are assumed to be iid $N(0; \sigma_i)$ random variables, $i = 1; 2; 3; 4$: Notice that α_4 in \mathcal{H}_4 corresponds to α_1 in the version of \mathcal{H}_4 given in Section 2.1, for example. In general, then, we are assuming that there is a deterministic trend in the time series under investigation, so that $S_{T;R}^a(\alpha)$ and $S_{T;R}^b(\alpha)$ do not need to be calculated, and $S_{T;R}(\alpha)$ is thus used throughout. In order to consider parameterizations which are illustrative of the types of DGPs observed in reality, we calibrate the

models using quarterly real and seasonally adjusted U.S. M2 for the period 1970:1-1994:1.⁸ This particular series is the money variable examined in King, Plosser, Stock and Watson (1991) and later updated by Corradi, Swanson and White (2000). After fixing $\frac{1}{2} = 0.75$, the following estimates were obtained,

$$\mathbf{X}_t = 73.31 + 8.84t + 0.75\mathbf{X}_{t-1}; \mathbf{\beta}_1 = 32.56,$$

$$\mathbf{X}_t = 30.78 + \mathbf{X}_{t-1}; \mathbf{\beta}_2 = 17.27$$

$$\log \mathbf{X}_t = 1.621 + 0.0050t + 0.75 \log \mathbf{X}_{t-1}; \mathbf{\beta}_3 = 0.0205, \text{ and}$$

$$\log \mathbf{X}_t = 0.0188 + \log \mathbf{X}_{t-1}; \mathbf{\beta}_4 = 0.0099.⁹$$

Using these estimated models as our “benchmark” models, data are generated according to DGPs with: $\mathbf{\beta}_1 = 75$, $\mathbf{\beta}_2 = (20; 30; 40; 50)$, $\mathbf{\beta}_3 = 2$, $\mathbf{\beta}_4 = (0.010; 0.015; 0.020; 0.025)$, $\mathbf{\beta}_1 = (5; 10; 15; 20)$, $\mathbf{\beta}_2 = (0.003; 0.004; 0.005; 0.006)$, and $\mathbf{\beta}_i$, $i = 1; 2; 3; 4$ is set equal to its estimated value. Samples of $T = 100, 200, 300, 400$, and 500 observations were simulated. Also, we set $\mathbf{R} = (T^{0.50}; T^{0.75}; T^{0.90}; T^{0.95})$. The range of \mathbf{u} is $-1.0 \leq \mathbf{u} \leq 1.0$, and 100 statistics for 100 increments within this range were calculated.¹⁰ All simulations are based on 500x500 Monte Carlo trials, where the first 500 corresponds to the number of Monte Carlo iterations, and the second 500 corresponds to the number of different \mathbf{u}_i , $i = 1; \dots; \mathbf{R}$ vectors that are drawn. (Put another way, for each new \mathbf{u} vector, a new statistic is calculated and inference based on that statistic is carried out. For each draw of the DGP, this is repeated 500 times.) Rejection frequencies based on these DGPs and a 5% nominal level are graphically depicted in Figures 1 and 2. The different plots in the figures correspond to different parameterizations, and are labelled as follows:

⁸Results based on models parameterized using U.S. GDP as a benchmark were also tabulated, and are available upon request.

⁹The above models are meant to be used only as benchmark DGPs, and are not necessarily indicative of what might be viewed as the “best” univariate linear time series model for money. In addition, if one of these models were assumed to be “true”, then the others would necessarily be misspecified. Further, there is likely a structural break in the money data being examined (see e.g. Swanson (1998)), so that all of the models may be rather inaccurate. These last two issues, while important for empirical analysis of the money data have no impact on our Monte Carlo analysis, however, as we are simply using the above parameterizations as given baseline models in our experiments (e.g. whether the parameters are consistent or not has no implications for the Monte Carlo experiments, per se). Please see the subsequent section for a more detailed discussion of the empirical properties of the money data used here.

¹⁰Various ranges and increments for \mathbf{u} were examined, including ranges for \mathbf{u} between -100 and 100. Results were found to be robust to the choice of \mathbf{U} and the number of increments.

Rejection Frequencies When Data Are Generated According to Log DGPs (Empirical Size)

DGP-S1: $\theta_4 = 0.010$, DGP-S2: $\theta_4 = 0.015$, DGP-S3: $\theta_4 = 0.020$, DGP-S4: $\theta_4 = 0.025$, DGP-S5: $\theta_3 = 2$, $\pm_2 = 0.003$, DGP-S6: $\theta_3 = 2$, $\pm_2 = 0.004$; DGP-S7: $\theta_3 = 2$, $\pm_2 = 0.005$, DGP-S8: $\theta_3 = 2$, $\pm_2 = 0.006$:

Rejection Frequencies When Data Are Generated According to Levels DGPs (Empirical Power)

DGP-P1: $\theta_2 = 20$, DGP-P2: $\theta_2 = 30$, DGP-P3: $\theta_2 = 40$, DGP-P4: $\theta_2 = 50$, DGP-P5: $\theta_1 = 75$, $\pm_1 = 5$, DGP-P6: $\theta_1 = 75$, $\pm_1 = 10$; DGP-P7: $\theta_1 = 75$, $\pm_1 = 15$, DGP-P8: $\theta_1 = 75$, $\pm_1 = 20$:

Turning to the results, recall first that the graphs denoted by DGP-S3 (DGP-S8) and DGP-P2 (DGP-P6) correspond most closely to the estimated models (i.e. to what we have termed our “benchmark” models), and in these cases, empirical rejection frequencies are close to the level of the test when data are generated under the null (DGP-S3 and DGP-S8), while rejection frequencies are above 0.60 for all values of \mathbf{R} except $\mathbf{R} = \mathbf{T}^{0.5}$, when data are generated under the alternative (DGP-P2 and DGP-P6). As expected, empirical power improves as we move from DGP-P1 to DGP-P4, for example, because the trend parameter (i.e. \pm_1 for \mathbf{H}_1 or θ_2 for \mathbf{H}_2) increases. The same argument can be made when viewing DGP-P5 - DGP-P8 in Figure 2. Correspondingly, increasing either \pm_2 or θ_4 results in improved empirical size, as evidenced by moving from DGP-S1 to DGP-S4, for example. Interestingly, empirical size is close to nominal in all cases, as long as samples of around 300 or more observations are used. The same can be said of empirical power, as rejection frequencies are generally above 0.80 in all cases except $\mathbf{R} = \mathbf{T}^{0.5}$, when samples of around 300 or more observations are used. The trade-off between smaller and bigger \mathbf{R} is also as expected - increasing \mathbf{R} results in worse empirical size and better empirical power. In summary, while our experiments are rather limited in scope, we have some evidence that the proposed test may be useful, even for samples of as few as 300 observations. However, empirical size/power trade-offs are very pronounced for smaller samples.

5 Empirical Illustration

In keeping with the Monte Carlo experiments reported on in the previous section, we now consider the quarterly U.S. data set examined by King, Plosser, Stock and Watson (KPSW: 1991), and updated in Corradi, Swanson and White (2000). In particular, the $\mathbf{S}_{\mathbf{T};\mathbf{R}}(!)$ test is carried out for four series, including: consumption, investment, money, and output. Note that variables of the

type examined here are all clearly upward trending, as documented in Stock and Watson (1989), for example, thus supporting our use of this particular version of the data transformation test. Also, note that the variables are constructed as in KPSW.¹¹ Results for a variety of values of \mathbf{R} , as well as for two different sub-samples, are given in Table 1.

Table 1: Empirical Illustration: The King-Plosser-Stock-Watson Data Set ^a

Series	Data Transformation Statistics				
	$\mathbf{R} = \mathbf{T}^{0:50}$	$\mathbf{R} = \mathbf{T}^{0:75}$	$\mathbf{R} = \mathbf{T}^{0:90}$	$\mathbf{R} = \mathbf{T}^{0:95}$	$\mathbf{R} = \mathbf{T}^{0:99}$
Panel A: Sample: 1947 quarter 1 - 1994 quarter 1					
Consumption	1.21	1.65	2.79	3.58	4.24
Investment	2.14	5.82	11.9	15.8	19.5
Money	2.59	7.73	16.2	21.4	26.4
Output	1.07	0.90	0.95	1.07	1.03
Panel B: Sample: 1970 quarter 1 - 1994 quarter 1					
Consumption	1.27	1.05	1.14	1.32	1.35
Investment	1.32	1.41	1.90	2.35	2.58
Money	1.80	3.65	6.35	7.96	9.41
Output	1.41	1.87	2.82	3.52	4.00

^a Entries in the table are $S_{\mathbf{T};\mathbf{R}}(!)$ statistics calculated as discussed above, and are distributed as \hat{A}_1^2 random variables so that 1% and 5% critical values are 6.63 and 3.84, respectively. Data are quarterly and correspond to those series constructed and examined by King, Plosser, Stock and Watson (1991), except that the data have been updated through 1994, as discussed in Corradi, Swanson and White (2000).

A number of conclusions can be made based on these results. First, consumption and output are best modelled in logs, a result that agrees in large part with previous empirical practice (see e.g. Engle and Granger (1987) and Diebold and Senhadji (1996)). Second, the evidence on investment is mixed. For the longer sub-sample, the statistics based on $\mathbf{R} = \mathbf{T}^{0:5}$ supports logs, while the statistics based on different choices of \mathbf{R} support levels. However, we know that the power of these tests is rather low for $\mathbf{R} = \mathbf{T}^{0:5}$: For this reason, and given that there is always a possibility of structural breaks (and hence poor test performance) among economic variables, we also constructed test statistics for the smaller sub-sample reported on in Panel B of the table. Notice that in this case, the null hypothesis of a loglinear DGP for investment is never rejected, regardless of the value of \mathbf{R} (the maximum value of the statistic is 2.58 and the 5% critical value is 3.84). Thus, although the evidence is somewhat mixed, it appears that investment is better modelled in logs,

¹¹Using citibase mnemonics, the series are constructed as follows: consumption=gccq/p; investment=gifq/p; money=fm2/p; output=(gdpq-ggeq)/p, with p=p16*1000000, p16=U.S. population, gccq=real consumption expenditures, fm2=nominal seasonally adjusted M2 stock, gdpq=real GDP, and ggeq=real government expenditures on goods and services. Thus, all series are per capita.

particularly if more recent data are being modelled. Third, the evidence on money is mixed. In both sub-samples, findings are dependent upon the choice of \mathbf{R} : Again, one reason for this may be the presence of a structural break. Indeed, in the early 1980's (prior to 1984) the federal reserve bank experimented with policy aimed at targeting the money stock. In addition, at about the same time, there was an apparent structural break in the money stock due to the introduction of interest bearing checking accounts and due to a surge in credit card usage, for example.¹² For these reasons, we also constructed statistics analogous to those reported in the table for a sub-sample beginning in 1984. For this sub-sample, the statistics for money are (2.04, 3.67, 6.19, 7.49, 8.42), for the various values of \mathbf{R} reported on in the table. Note that although there is now stronger evidence than before for modelling money in logs, the evidence is still mixed. Thus, no definite choice among logs and levels is provided by the test when modelling money. Overall, though, this illustration supports the common practice in empirical macroeconomics of logarithmic data transformation prior to unit root testing.

6 Concluding Remarks

Unit root and stationarity tests are severely biased, both in small and in large sample, in the presence of incorrect data transformation. In this paper we have proposed a simple test, based on the combined use of a randomization procedure and sample conditioning, for choosing between linearity in logs and linearity in levels, in the presence of deterministic and/or stochastic trends. For any sample which is a realization of a DGP under the null hypothesis (i.e. a log DGP), the statistic has a $\hat{\mathbf{A}}^2$ limiting distribution, while for any sample which is a realization of a DGP under the alternative (i.e. a level DGP) the statistic diverges. Once we have chosen the correct the data transformation, we remain with the standard problem of testing for a unit root. A Monte Carlo exercise is used to examine the finite sample behavior of the suggested testing procedure, and our findings are rather encouraging for samples of at least 300 observations. In addition, an empirical illustration based on the King, Plosser, Stock and Watson (1991) data set is given, and evidence of preference for the loglinear model is provided.

¹²See Clements and Hendry (1999a,b) for a detailed discussion of forecasting failure in the presence of structural breaks in economic series.

7 Appendix

Proof of Theorem 1: (i) First note that conditional on the sample, $\mathbf{8i}; \mathbf{V}_{\gg;T}^i \gg \mathbf{N}^{\mu} \mathbf{0}; \frac{1}{T} \sum_{t=1}^T \frac{\mathbb{P}_T^3 \frac{\mathbb{C}X_t(!)}{\mathbb{C}X_1(!)} \hat{2}^{\mathfrak{N}}}{\mathbb{C}X_1(!)}$. Let $\Omega^+ = \mathbf{f}! : \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{\mathbb{P}_T^3 \frac{\mathbb{C}X_t(!)}{\mathbb{C}X_1(!)} \hat{2}^{\mathfrak{N}}}{\mathbb{C}X_1(!)} = \mathbf{1} \mathbf{g}$: We begin by showing that $\mathbf{P}(\Omega^+) = 1$: Under DGP \mathbf{H}_3 ; $\Delta \mathbf{X}_t = \mathbf{X}_{t-1} \exp(\pm_1 \sum_{j=0}^{t-1} \mathfrak{W}^j + \sum_{j=0}^{t-1} \mathfrak{W}^j ("_{2;t_{ij}} \mathbf{i} "_{2;t_{ij}-1}))$; where $\mathbf{X}_{t-1} = \exp(\log \mathbf{X}_0 + \sum_{j=0}^{t-1} \mathfrak{W}^j + \sum_{j=0}^{t-1} \mathfrak{W}^j (t_{ij}) + \sum_{j=0}^{t-1} \mathfrak{W}^j "_{2;t_{ij}-1})$: Under DGP \mathbf{H}_4 ; $\Delta \mathbf{X}_t = \mathbf{X}_{t-1} \exp(\pm_1 + "_{2;t})$; where $\mathbf{X}_{t-1} = \exp(\log \mathbf{X}_0 + \sum_{j=0}^{t-1} \mathfrak{W}^j + \sum_{j=0}^{t-1} \mathfrak{W}^j "_{2;j})$: The functional law of the iterated logarithm for strong mixing processes (e.g. Berger Theorem 3.1, 1990) states that $\limsup_{t \rightarrow \infty} \frac{1}{\sqrt{2t \log \log t}} \sum_{j=1}^t \mathfrak{W}^j = \mathbf{O}_{a.s.}(1)$: The deterministic trend component is then the dominant term in both DGPs. It follows that $\frac{1}{T} \sum_{t=1}^T \frac{\mathbb{P}_T^3 \frac{\mathbb{C}X_t(!)}{\mathbb{C}X_1(!)} \hat{2}^{\mathfrak{N}}}{\mathbb{C}X_1(!)} \xrightarrow{a.s.} \mathbf{1}$; at a geometric rate. Thus, $\mathbf{P}(\Omega^+) = 1$: We now proceed conditionally on the sample, and with the notation $\mathbf{a:s: i}!$; we mean conditionally on $\mathbf{!} \geq \Omega^+$: Hereafter, let $\mathbf{U} = [\underline{\mathbf{u}}; \bar{\mathbf{u}}]$: We first need to show that for $\mathbf{8u} \geq \mathbf{U}$;

$$\mathbf{P}^{\mathfrak{a}}(\mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u}) = \frac{1}{2} + \mathbf{O}(T^{-1/2}); \quad (5)$$

where the $\mathbf{O}(T^{-1/2})$ term holds uniformly in \mathbf{i} and \mathbf{u} ; $\mathbf{a:s: i}!$: Suppose $\mathbf{u} > 0$; then

$$\mathbf{P}^{\mathfrak{a}}(\mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u}) = \mathbf{P}^{\mathfrak{a}}(\mathbf{V}_{\gg;T}^i(!) \cdot 0) + \mathbf{P}^{\mathfrak{a}}(0 \cdot \mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u}) \mathbf{a:s: i}!$$

As \mathfrak{v}_i is a zero mean normal, $\mathbf{P}^{\mathfrak{a}}(\mathbf{V}_{\gg;T}^i(!) \cdot 0) = \frac{1}{2}$: Therefore, it suffices to show that $\mathbf{P}^{\mathfrak{a}}(0 \cdot \mathbf{V}_{\gg;T}^i \cdot \mathbf{u}) = \mathbf{O}(T^{-1/2})$; uniformly in \mathbf{u} and \mathbf{i} ; $\mathbf{a:s: i}!$: Now,

$$\begin{aligned} \mathbf{P}^{\mathfrak{a}}(0 \cdot \mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u}) &= \frac{1}{\frac{1}{T} \sum_{t=1}^T \frac{\mathbb{P}_T^3 \frac{\mathbb{C}X_t(!)}{\mathbb{C}X_1(!)} \hat{2}^{\mathfrak{N}}}{\mathbb{C}X_1(!)}} \int_0^{\bar{\mathbf{u}}} \exp \left(-\frac{\tilde{\mathbf{A}}}{2} \mathbf{x}^2 = \frac{1}{T} \sum_{t=1}^T \frac{\mathbb{P}_T^3 \frac{\mathbb{C}X_t(!)}{\mathbb{C}X_1(!)} \hat{2}^{\mathfrak{N}}}{\mathbb{C}X_1(!)} \right) d\mathbf{x} \\ &= \mathbf{O}(T^{-1/2}); \text{ uniformly in } \mathbf{i}; \mathbf{a:s: i}!; \end{aligned} \quad (6)$$

as $\sup_{\mathbf{u} \geq \mathbf{U}} \int_0^{\bar{\mathbf{u}}} \exp \left(-\frac{\tilde{\mathbf{A}}}{2} \mathbf{x}^2 = \frac{1}{T} \sum_{t=1}^T \frac{\mathbb{P}_T^3 \frac{\mathbb{C}X_t(!)}{\mathbb{C}X_1(!)} \hat{2}^{\mathfrak{N}}}{\mathbb{C}X_1(!)} \right) d\mathbf{x} \leq \bar{\mathbf{u}}$; $\mathbf{a:s: i}!$; and $\frac{1}{T} \sum_{t=1}^T \frac{\mathbb{P}_T^3 \frac{\mathbb{C}X_t(!)}{\mathbb{C}X_1(!)} \hat{2}^{\mathfrak{N}}}{\mathbb{C}X_1(!)}$ diverges at a faster rate than T : A similar argument applies to the case of $\mathbf{u} < 0$: Hereafter, $\mathbf{E}^{\mathfrak{a}}$ denotes the expectation with respect to the probability measure $\mathbf{P}^{\mathfrak{a}}$: Now, for any given $\mathbf{u} \geq \mathbf{U}$;

$$\begin{aligned} \frac{1}{\mathbf{P}(\bar{\mathbf{R}})} \sum_{i=1}^{\mathfrak{R}} \mathbf{1} \mathbf{f} \mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u} \mathbf{g} \mathbf{i} \frac{1}{2} &= \frac{1}{\mathbf{P}(\bar{\mathbf{R}})} \sum_{i=1}^{\mathfrak{R}} \mathbf{1} \mathbf{f} \mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u} \mathbf{g} \mathbf{i} \mathbf{E}^{\mathfrak{a}}(\mathbf{1} \mathbf{f} \mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u} \mathbf{g}) \\ &\quad + \frac{1}{\mathbf{P}(\bar{\mathbf{R}})} \mathbf{E}^{\mathfrak{a}}(\mathbf{1} \mathbf{f} \mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u} \mathbf{g}) \mathbf{i} \frac{1}{2} : \end{aligned} \quad (7)$$

Note that $\mathbf{E}^{\mathfrak{a}}(\mathbf{1} \mathbf{f} \mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u} \mathbf{g}) = \mathbf{P}^{\mathfrak{a}}(\mathbf{V}_{\gg;T}^i(!) \cdot \mathbf{u}) = \frac{1}{2} + \mathbf{O}(T^{-1/2})$; where the $\mathbf{O}(T^{-1/2})$ term holds uniformly in \mathbf{i} and \mathbf{u} ; $\mathbf{a:s: i}!$. As \mathbf{R} grows at a rate slower than T ; the last term on the RHS of

(7) approaches zero, $a:s: j \rightarrow \infty$: Recall that Var^a denotes the variance with respect the probability measure P^a : Now, as $V_{\gg,T}^i(!)$ is independent of $V_{\gg,T}^j(!)$; $8i = j$; $a:s: j \rightarrow \infty$; and recalling (5),

$$\begin{aligned} \text{Var}^a \left(\frac{1}{R} \sum_{i=1}^R 1fV_{\gg,T}^i(!) \cdot ug \right) &= \frac{1}{R} \sum_{i=1}^R (E^a(1fV_{\gg,T}^i(!) \cdot ug^2) - (E^a(1fV_{\gg,T}^i(!) \cdot ug))^2) \\ &= \frac{1}{R} \sum_{i=1}^R E^a(1fV_{\gg,T}^i(!) \cdot ug) - \frac{1}{4} + O(T^{-1}) \\ &= \frac{1}{2} + O(T^{-1/2}) - \frac{1}{4} + O(T^{-1}) = \frac{1}{4} + O(T^{-1/2}); \end{aligned}$$

uniformly in i and u ; $a:s: j \rightarrow \infty$. By noting that $1fV_{\gg,T}^i(!) \cdot ug$; $i = 1, \dots, R$ and $R = T^a$; $0 < a < 1$; is an independent triangular array, by the central limit for independent triangular arrays (see e.g. Davidson (2000, p.52)), for all $u \in U$;

$$\frac{1}{R} \sum_{i=1}^R 1fV_{\gg,T}^i(!) \cdot ug \xrightarrow{d} N(0, 1/4):$$

We now need to show that the convergence above holds uniformly in u : That is, we need to show that

$$\frac{1}{R} \sum_{i=1}^R 1fV_{\gg,T}^i(!) \cdot ug - \frac{1}{R} \sum_{i=1}^R 1fu \cdot V_{\gg,T}^i(!) \cdot u^0g = o_{P^a}(1);$$

with the $o_{P^a}(1)$ term independent of u and u^0 : Without loss of generality, let $u < u^0$: Then,

$$\frac{1}{R} \sum_{i=1}^R 1fV_{\gg,T}^i(!) \cdot ug - 1fV_{\gg,T}^i(!) \cdot u^0g = \frac{1}{R} \sum_{i=1}^R 1fu \cdot V_{\gg,T}^i(!) \cdot u^0g :$$

Now,

$$\begin{aligned} \sup_{u; u^0 \in U} P^a \left(\left| \frac{1}{R} \sum_{i=1}^R 1fu \cdot V_{\gg,T}^i(!) \cdot u^0g \right| > \frac{1}{2} \right) &\leq \frac{1}{2^2} \sup_{u; u^0 \in U} E^a \left(\frac{1}{R} \sum_{i=1}^R 1fu \cdot V_{\gg,T}^i(!) \cdot u^0g \right)^2 \\ &= \frac{1}{2^2} \frac{1}{R} \sum_{i=1}^R \sup_{u; u^0 \in U} E^a \left(1fu \cdot V_{\gg,T}^i(!) \cdot u^0g \right)^2 = \frac{1}{2^2} \sup_i P^a(u \cdot V_{\gg,T}^i \cdot u) = O(T^{-1/2}); \end{aligned}$$

$a:s: j \rightarrow \infty$; because of (6). The desired result then follows.

(ii) Let $\Omega_A = \{! : \frac{1}{T} \sum_{t=1}^T \frac{\phi X_t(!)}{\phi X_1(!)} \leq M; 0 < M < 1\}$. We begin by showing that $P(\Omega_A) = 1$: Now, $\Delta X_t = \pm 0 \sum_{j=0}^{t-1} \frac{1}{j} + \sum_{j=0}^{t-1} \frac{1}{j} \frac{1}{j} (!_{1:t,j} - !_{1:t,j-1})$; under DGP H_1 ; and $\Delta X_t = \pm 0 + !_{1:t}$; under H_2 : Given A1, it follows by the strong law of large numbers that $\frac{1}{T} \sum_{t=1}^T \frac{\phi X_t}{\phi X_1} \xrightarrow{a.s.} M$; and so $P(\Omega_A) = 1$: (Hereafter, with the notation $a:s: j \rightarrow \infty$; we mean for all $! \in \Omega_A$.) From the previous statements, it follows that $V_{\gg,T}^i(!)$ is a zero mean normal random variable with variance equal to

$\frac{1}{T} \sum_{t=1}^T \frac{1}{\sqrt{\Delta X_1(!)}} \left(\frac{\Delta X_t(!)}{\Delta X_1(!)} \right)^2$; so that $V_{\gg;T}^i(!) \xrightarrow{d} N(0; M)$; as $s \rightarrow \infty$; as $T \rightarrow \infty$; and 8i: Let $F(u)$ be the cumulative distribution function (CDF) of a $N(0; M)$; evaluated at $u \in \mathbb{U}$: Then,

$$\frac{2}{\sqrt{R}} \sum_{i=1}^R (1fV_{\gg;T}^i(!) \cdot ug) \cdot \frac{1}{2} = \frac{2}{\sqrt{R}} \sum_{i=1}^R (1fV_{\gg;T}^i(!) \cdot ug) \cdot F(u) + 2\sqrt{R}(F(u) \cdot \frac{1}{2}); \quad (8)$$

As $F(u) = \frac{1}{2}$ when $u = 0$; the second term on the RHS of (8) diverges to $+\infty$ or $-\infty$ at rate \sqrt{R} ; as $s \rightarrow \infty$; for all $u = 0$: In addition, the first term on the RHS of (8) is bounded in probability, as can be shown by noting that,

$$\frac{2}{\sqrt{R}} \sum_{i=1}^R (1fV_{\gg;T}^i(!) \cdot ug) \cdot F(u) = \frac{2}{\sqrt{R}} \sum_{i=1}^R (1fV_{\gg;T}^i(!) \cdot ug) \cdot F_T(u) + 2\sqrt{R}(F_T(u) \cdot F(u));$$

where $F_T(u) = P^\pi(V_{\gg;T}^i(!) \cdot u)$: As $V_{\gg;T}^i$ has finite variance and is independent i , the Berry-Essen theorem (e.g. Davidson (1994) p.408) can be applied, yielding that,

$$\sup_{u \in \mathbb{U}} (F_T(u) \cdot F(u)) = O(T^{-1/2}); \text{ as } s \rightarrow \infty:$$

In addition, as $R=T \rightarrow \infty$; $\sup_{u \in \mathbb{U}} \sqrt{R}(F_T(u) \cdot F(u)) \rightarrow 0$; as $s \rightarrow \infty$: Now, $E^\pi(1fV_{\gg;T}^i(!) \cdot ug) = \int_{-\infty}^u dF_T(s) = F_T(u)$; and $Var^\pi(1fV_{\gg;T}^i(!) \cdot ug) = F_T(u)(1 - F_T(u))$: Thus, as $R=T \rightarrow \infty$; $\frac{2}{\sqrt{R}} \sum_{i=1}^R (1fV_{\gg;T}^i(!) \cdot ug) \cdot F_T(u) \xrightarrow{d} N(0; 4F(u)(1 - F(u)))$; as $s \rightarrow \infty$; as $F_T(u) \rightarrow F(u)$ when $T \rightarrow \infty$: Also, as $\frac{1}{\sqrt{R}} \sum_{i=1}^R (1fV_{\gg;T}^i(!) \cdot ug)$ is stochastic equicontinuous on \mathbb{U} ; it follows that $\sup_{u \in \mathbb{U}} \frac{1}{\sqrt{R}} \sum_{i=1}^R (1fV_{\gg;T}^i(!) \cdot ug)$ weakly converges to the supremum of a Gaussian process. Thus, the LHS of (8) diverges in probability at rate \sqrt{R} ; 8! $2 \Omega_A$; with $Pr(\Omega_A) = 1$:

Proof of Theorem 2: (i) Under DGP H_4 ; $\pm_1 = 0$ and $\Delta X_t = X_{t+1} \exp(\epsilon_{2;t})$; where $X_{t+1} = \exp(\log X_0 + \sum_{j=1}^t \epsilon_{2;j})$: Let

$$\Omega_1 : f! : \frac{1}{T} \sum_{t=1}^T \mu \frac{\Delta X_t(!)}{\Delta X_1(!)} \xrightarrow{d} N(0; 1);$$

and

$$\Omega_2 : f! : \frac{1}{T} \sum_{t=1}^T \mu \frac{\Delta X_t(!)}{\Delta X_1(!)} \xrightarrow{d} N(0; 0);$$

We begin by establishing that $P(\Omega_1 \cap \Omega_2) = 1$: This can be done by first showing that $Pr(\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{j=1}^t \epsilon_{2;j} = 1) = 1$: Given A1, the strong invariance principle for stationary α -mixing processes (e.g. Eberlain (1986), Theorem 2) ensures that,

$$r = \frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor T \rfloor} \mathbf{X}_j = \mathbb{W}(r) + O_{a.s.}(T^{-\mu} \log \log T); \text{ for } 0 < \mu < 1/2;$$

where $r \in (0, 1]$; $\mathbb{W}^2 = E(\mathbb{W}_{2,t}^2)$; and \mathbf{W} is a standard Brownian motion process. Now define,

$$\Psi = f(t; \cdot) \in [0, 1] \times \Omega : \mathbf{W}(t; \cdot) = 0g;$$

and $\mathbb{B} \in \Omega$; define,

$$\Psi(\cdot) = f(t \in [0, 1]) : \mathbf{W}(t; \cdot) = 0g;$$

From Theorem 2.9.6 in Karatzas and Shreve (1991), it follows that $\Psi(\cdot)$ has zero Lebesgue measure, $\mathbb{B} \in \Omega^c$; where $P(\Omega^c) = 1$. Thus, it also follows that as $t \rightarrow 1$; $\frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor T \rfloor} \mathbf{X}_j \xrightarrow{a.s.} 1$, at rate $T^{-\mu}$; $0 < \mu < 1/2$. This implies that \mathbb{B} for which $\frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor T \rfloor} \mathbf{X}_j(\cdot) \rightarrow 1$; $\Delta \mathbf{X}_t(\cdot)^2 \rightarrow 1$; and \mathbb{B} for which $\frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor T \rfloor} \mathbf{X}_j(\cdot) \rightarrow 1$; $\Delta \mathbf{X}_t(\cdot)^2 \rightarrow 0$. Now, $\Pr(\Delta \mathbf{X}_1 = 0) = 0$; and given the moment conditions in A1, $\frac{1}{\sqrt{T}} \Delta \mathbf{X}_1 \xrightarrow{a.s.} 0$; Thus, \mathbb{B} for which $\Delta \mathbf{X}_t(\cdot)^2 \rightarrow 1$; we also have that $(\Delta \mathbf{X}_t(\cdot) - \Delta \mathbf{X}_1(\cdot))^2 \rightarrow 1$; and \mathbb{B} for which $\Delta \mathbf{X}_t(\cdot)^2 \rightarrow 0$; $(\Delta \mathbf{X}_t(\cdot) - \Delta \mathbf{X}_1(\cdot))^2 \rightarrow 0$; both at a geometric rate. It follows that $\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\mathbb{E} \mathbf{X}_t(\cdot)^2}{\mathbb{E} \mathbf{X}_1(\cdot)^2} \rightarrow 1$ or $\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\mathbb{E} \mathbf{X}_t(\cdot)^2}{\mathbb{E} \mathbf{X}_1(\cdot)^2} \rightarrow 0$, \mathbb{B} ; also at a geometric rate. Thus, $P(\Omega_1 \cap \Omega_2) = 1$: It remains to establish that as $T; R \rightarrow 1$ and $R=T \rightarrow 0$; (a) $\min(\mathbf{S}_{T;R}^a(\cdot); \mathbf{S}_{T;R}^b(\cdot)) = \mathbf{S}_{T;R}^a(\cdot)$ and $\mathbf{S}_{T;R}^a(\cdot) \xrightarrow{P} \hat{\mathbf{A}}_1^2$ a.s.; $\mathbb{B} \in \Omega_1$; and (b) $\min(\mathbf{S}_{T;R}^a(\cdot); \mathbf{S}_{T;R}^b(\cdot)) = \mathbf{S}_{T;R}^b(\cdot)$ and $\mathbf{S}_{T;R}^b(\cdot) \xrightarrow{P} 0$ a.s.; $\mathbb{B} \in \Omega_2$; (with the notation \xrightarrow{P} we mean convergence in probability according to P ; conditionally on the sample).

(a) That $\mathbf{S}_{T;R}^a(\cdot) \xrightarrow{P} \hat{\mathbf{A}}_1^2$ a.s.; $\mathbb{B} \in \Omega_1$ follows directly by the same arguments used in the proof of Theorem 1(i). Now,

$$\begin{aligned} \mathbf{S}_{T;R}^b(\cdot) &= \int_{u^+}^Z \left(\frac{1}{\sqrt{R}} \sum_{i=1}^R ((1 - \mathbf{V}_{\cdot;T}^i(\cdot)) \cdot u - \frac{1}{2}) - \frac{1}{2} \right)^2 \mathbb{W}(u) du \\ &= \mathbf{S}_{T;R}^a(\cdot) + \frac{1}{4} \frac{1}{\sqrt{R}} \int_{u^+}^Z \left(\frac{1}{\sqrt{R}} \sum_{i=1}^R (1 - \mathbf{V}_{\cdot;T}^i(\cdot)) \cdot u - \frac{1}{2} \right) \mathbb{W}(u) du; \end{aligned}$$

so that $\mathbf{S}_{T;R}^b(\cdot)$ diverges at rate $\frac{1}{\sqrt{R}}$:

(b) Recall that $\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\mathbb{E} \mathbf{X}_t(\cdot)^2}{\mathbb{E} \mathbf{X}_1(\cdot)^2} \rightarrow 0$; $\mathbb{B} \in \Omega_2$: As $\mathbf{V}_{\cdot;T}^i(\cdot)$ is a zero mean normal with variance equal to $\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\mathbb{E} \mathbf{X}_t(\cdot)^2}{\mathbb{E} \mathbf{X}_1(\cdot)^2}$; it follows that $\mathbf{V}_{\cdot;T}^i(\cdot) \xrightarrow{P} 0$ a.s.; $\mathbb{B} \in \Omega_2$; \mathbb{B} : Furthermore,

$\frac{1}{T} \sum_{t=1}^T \left(\frac{\langle X_t(\cdot) \rangle^2}{\langle X_1(\cdot) \rangle} \right) \rightarrow 0$ at an exponential rate and $V_{\gg;T}^i \xrightarrow{P} 0$; $\| \cdot \|_2 \leq \Omega_2$; at the same rate. Thus, $\|u\|_2 \leq U^+$; $\langle V_{\gg;T}^i(\cdot) \cdot u \rangle \xrightarrow{P} 1$; $\|a\|_2 \leq U^+$; $\| \cdot \|_2 \leq \Omega_2$; at an exponential rate as $T \rightarrow \infty$. It follows that $\|u\|_2 \leq U^+$; as $T; R \rightarrow \infty$; $R=T \rightarrow 0$; $\frac{1}{R} \sum_{i=1}^R (\langle V_{\gg;T}^i \cdot u \rangle - 1) \xrightarrow{P} 0$; and so $S_{T;R}^b(\cdot) \xrightarrow{P} 0$; $\|a\|_2 \leq U^+$; $\| \cdot \|_2 \leq \Omega_2$. Also, note that

$$S_{T;R}^a(\cdot) = S_{T;R}^b(\cdot) + \frac{1}{4} \frac{1}{R} \int_{U^+}^Z \left(\frac{1}{R} \sum_{i=1}^R (\langle V_{\gg;T}^i(\cdot) \cdot u \rangle - \frac{1}{2}) \right) \mathbb{1}(u) du;$$

so that $S_{T;R}^a(\cdot)$ diverges at rate $\frac{1}{R}$:

8 References

- Arino, M.A. and P.H. Franses, (2000), Forecasting the Level of Vector Autoregressions of Log Transformed Series, **International Journal of Forecasting**, 16, 111-116.
- Berger, E., (1990), An Almost Sure Invariance Principle for Stationary Ergodic Sequences of Banach Space Valued Random Variables, **Probability Theory and Related Fields**, 84, 161-201.
- Bickel, P.J. and K. Doksum, (1981), An Analysis of Transformations Revisited, **Journal of the American Statistical Association**, 76, 296-310.
- Chao, J.C., V. Corradi and N.R. Swanson, (2001), Data Transformation and Forecasting in Models with Unit Roots and Cointegration, **Annals of Economics and Finance**, 2, 59-76.
- Clements, M.P. and D.F. Hendry, (1999a), **Forecasting Economic Time Series: The Zeuthen Lectures on Economic Forecasting**, MIT Press, Cambridge.
- Clements, M.P. and D.F. Hendry, (1999b), On Winning Forecasting Competitions in Economics, **Spanish Economic Review**, 1, 123-160.
- Corradi, V., (1995), Nonlinear Transformation of Integrated Time Series: a Reconsideration, **Journal of Time Series Analysis**, 16, 539-550.
- Corradi, V. and N.R. Swanson, and H. White, (2000), Testing for Stationary Ergodicity and Co-movements Between Nonlinear Discrete Time Markov Processes, **Journal of Econometrics**, 96, 39-73.
- Corradi, V. and N.R. Swanson, (2001), A Consistent Test for Nonlinear Out of Sample Predictive Accuracy, Mimeo.
- Davidson, J., (1994), **Stochastic Limit Theory**, Oxford University Press, Oxford.
- Davidson, J., (2000), **Econometric Theory**, Blackwell Publishers, Oxford.
- De Bruin, P. and P.H. Franses, (1999), Forecasting Power Transformed Time Series, **Journal of Applied Statistics**, 27, 807-815.
- Diebold, F.X. and A.S. Senhadji, (1996), Deterministic versus Stochastic Trends in US GNP, Yet Again, **American Economic Review**, 86, 1291-1298.
- Dickey, D.A. and W.A. Fuller, (1979), Distribution of the Estimators for Autoregressive Time Series with a Unit Root, **Journal of the American Statistical Association**, 74, 427-431.
- Dufour, J.M. and J.F. Kiviet, (1996), Exact Tests for Structural Change in First-Order Dynamic Models, **Journal of Econometrics**, 70, 39-68.

- Eberlain, E., (1986), On Strong Invariance Principles under Dependence Assumptions, **Annals of Probability**, 14, 260-270.
- Engle, R.F. and C.W.J. Granger, (1987), Co-Integration and Error Correction: Representation, Estimation, and Testing, **Econometrica**, 55, 251-276.
- Franses, P.H. and G. Koop, (1998), On the Sensitivity of Unit Root Inference to Nonlinear Data Transformations, **Economics Letters**, 59, 7-15.
- Franses, P.H. and M. McAleer, (1998), Testing for Unit Roots and Nonlinear Transformations, **Journal of Time Series Analysis**, 19, 147-164.
- Granger, C.W.J. and J. Hallman, (1991), Nonlinear Transformations of Integrated Time Series, **Journal of Time Series Analysis**, 12, 207-224.
- Hansen, B.E., (1996), Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis, **Econometrica**, 64, 413-430.
- Inoue, A., (2001), Testing for Distributional Change in Time Series, **Econometric Theory**, 17, 156-187.
- Karatzas, J. and S.E. Shreve, (1991), **Brownian Motion and Stochastic Calculus**, Springer-Verlag, New York.
- King, R.G., C.I. Plosser, J.H. Stock and M. Watson, (1991), Stochastic Trends and Economic Fluctuations, **American Economic Review**, 81, 819-840.
- Kobayashi, M., (1994), Power of Tests for Nonlinear Transformation in Regression Analysis, **Econometric Theory**, 10, 357-371.
- Kobayashi, M. and M. McAleer, (1999a), Tests of Linear and Logarithmic Transformations for Integrated Processes, **Journal of the American Statistical Association**, 94, 860-868.
- Kobayashi, M. and M. McAleer, (1999b), Analytical Power Comparisons of Nested and Nonnested Tests for Linear and Loglinear Regression Models, **Econometric Theory**, 15, 99-113.
- Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin, (1992), Testing for the Null Hypothesis of Stationarity against the Alternative of a Unit Root, **Journal of Econometrics**, 54, 159-178.
- Long, J.B. and C.I. Plosser, (1982), Trends and Random Walks in Macroeconomic Time Series, **Journal of Political Economy**, 91, 39-69.
- Lütkepohl, H. and M.M. Burda, (1997), Modified Wald Tests under Nonregular Conditions, **Journal of Econometrics**, 78, 315-332.
- Park, J.Y. and P.C.B. Phillips, (1999), Asymptotics for Nonlinear Transformations of Integrated

Time Series, **Econometric Theory**, 15, 269-298.

Park, J.Y. and P.C.B. Phillips, (2001), Nonlinear Regression with Integrated Time Series, **Econometrica**, 69, 117-162.

Revuz, D. and M. Yor, (1990), **Continuous Martingales and Brownian Motion**, Springer and Verlag, Berlin.

Stock, J.H. and M.W. Watson, (1989), Interpreting the Evidence on Money-Income Causality, **Journal of Econometrics**, 40, 161-181.

Swanson, N.R., (1998), Money and Output Viewed Through a Rolling Window, **Journal of Monetary Economics**, 41, 455-474.

Figure 1: Monte Carlo Results for I(1) Data Generating Processes
5% Nominal Size Test Results

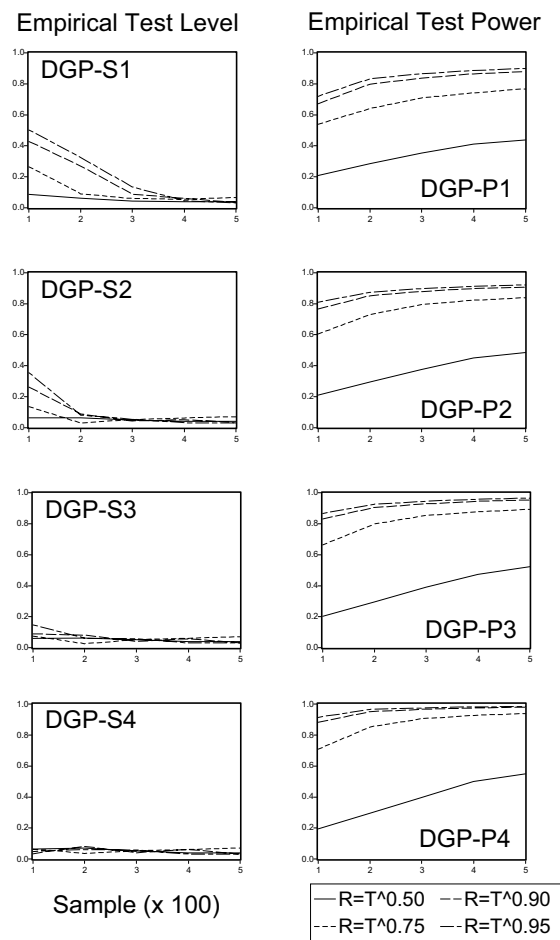


Figure 1:

Figure 2: Monte Carlo Results for I(0) Data Generating Processes
5% Nominal Size Test Results

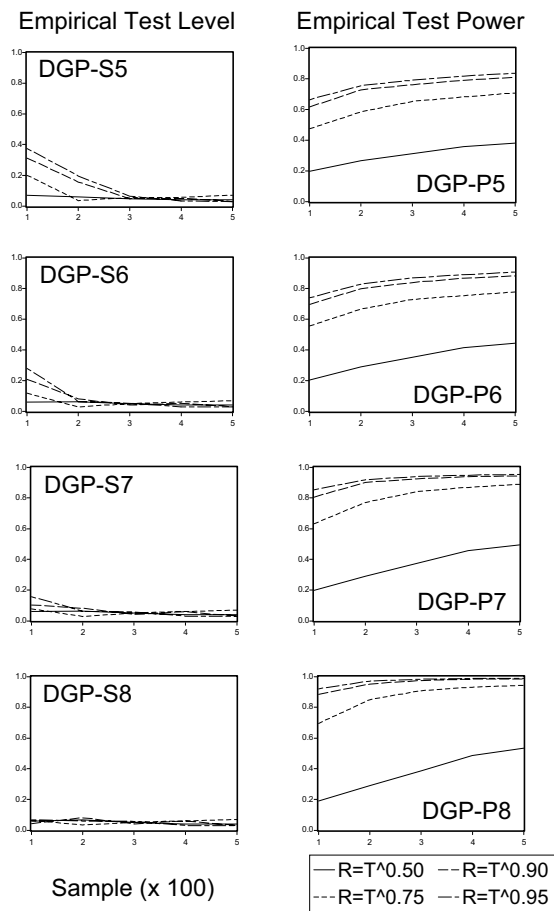


Figure 2: